

Exploring Hollywood's Appeal to the International Audience Through Foreign Actors

Sharon Chen
Stanford University
sharon19@stanford.edu

Da Eun Kim
Stanford University
daeunk@stanford.edu

ABSTRACT

There is no denying that Hollywood has an incredible international audience, but many of these countries also have their own domestic films being created. Given this, we proceed to conduct an exploratory analysis around the presence of foreign actors in Hollywood films and how these films perform in respective countries. Questions being explored include: relationship between presence of foreign actor in movie and movie's success in foreign actor's country, and factors that may contribute to the success of Hollywood movies in foreign countries.

Author Keywords

Hollywood; film; entertainment; cast; actors; data; analysis; visualization

INTRODUCTION

"Understanding international preferences is crucial as Hollywood and independent filmmakers reach for larger global shares. Worldwide ticket sales reached a record-breaking \$38.3 billion last year. More than 70% of the film industry's box office is generated overseas" [1]. It is evident that the international audience is incredibly important to a Hollywood film's success. So when events like a Hollywood blackout in China happen, this potentially pushes Hollywood to feature foreign actors in their movies [2].

However, amidst the popularity of these actors in their respective homeland, we're curious to further explore this relationship of the presence of a foreign actor in a Hollywood movie with its success in the box office domestically, internationally, and in that actor's homeland. Is it enough for the actor to be present in the movie for it to be significantly more successful abroad, or do they need a certain number of lines/screen time for there to be a significant difference?

In addition, we were particularly inspired by issues around gender and racial representation in Hollywood. There is much literature done on the gender divide in Hollywood (which we shall elaborate more on below), but we wanted to focus on foreign actors because there could be motivations other than representation for Hollywood to cast them in their movies.

RELATED WORK

There has not been much literature surrounding this topic of foreign actors in Hollywood movies. Instead, we opted to look for any visualizations relating to films and people associated with the films.

The Hollywood In\$ider is a visualization in the form of a scatter plot [3]. It provides a lot of options to toggle the y-axis from percentage of the budget recovered to opening weekend sales. However, it feels quite limiting in being constrained to the scatterplot, and it does not give us further information on foreign countries aside from the cumulative foreign gross. All in all, it gives us a good idea of what data we should be able to collect (since there is no consolidated source of data), and we later draw inspiration from this to also model box office data and a scatterplot showing different data.

The visualization we were able to find about films that concerned people associated with the films is "Film Dialogue" by The Pudding [4]. The Pudding is a website oriented around data journalism, creating visualizations that are interactive while intuitive to use. It also has a bit of text, much like a newspaper article would. This particular article presents visualizations surrounding film dialogue by gender in Disney movies, but we are not able to see its relationship with potentially box office success, which is the particular relationship we are interested in.

As a result, we are able to see that our question has not really been answered by any other previous literature and visualizations, but we would also noted a statement from the Pudding: "We didn't set out trying to prove anything, but rather compile real data. We framed it as a census rather than a study. So we Googled our way to 8,000 screenplays and matched each character's lines to an actor. From there, we compiled the number of words spoken by male and female characters across roughly 2,000 films, arguably the largest undertaking of script analysis, ever." From this, we get the sense that most of our time will be spent on data collection, and due to the massive amount of data we plan on collecting for our visualizations, it is best to frame it as a census much like The Pudding, and to allow our users to make their own interpretations and form their own questions from the data.

METHODS

Below we detail our process of prototyping our visualizations and collecting data. It is important to note that most of our time in the past half quarter collecting data!

Sketching Visualizations

When we began this project, we started with 2 big questions that we wanted to address with our visualizations:

1. Is there a significant relationship between the presence of a foreign actor in a Hollywood movie and its success in the foreign actor's nation?
2. Are there thresholds such as an actor's screen time/amount of lines that influence box office success?

Before we address those questions, we thought having a more generic visualization would be useful so that we can observe overall trends and comparisons of foreign actors. The general questions we have are:

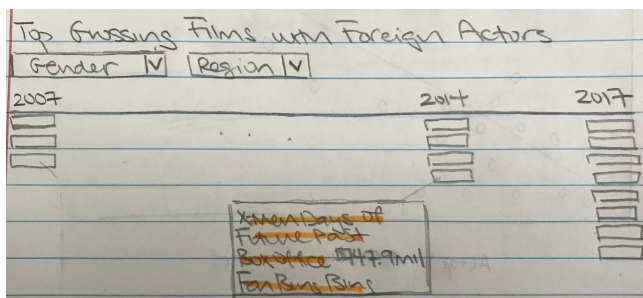
1. Where do the most foreign actors come from?

First, we must identify what "foreign" means. Because we are concerned about Hollywood in particular, we first made sure that the movies we are looking at had some US contribution in them. From there, we were able to define "foreign" by nationality. Thus, an actor would be considered foreign if their nationality is not the USA.

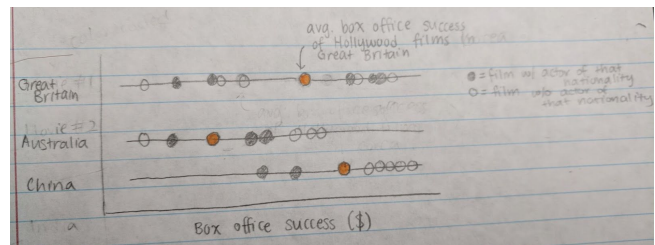
To address this question, we decided a heat map of the world would be the best visualization to show us geographically where foreign actors comes from.

2. Is there any change in the number of foreign actors over time?

We thought a time series-based histogram as seen in Picture 1 would be the best visualization so that we could also explore each individual movie. We wanted to make our visualizations interactive like The Pudding, so we thought adding filters around gender or continent would be useful as well.



Picture 1. Histogram of films with foreign actors in them from 2007-2017.



Picture 2. Dot plot addressing the question "Is there a significant relationship between the presence of a foreign actor in a Hollywood movie and its success in the foreign actor's nation?"

Now going back to our initial 2 questions, Pictures 2 and 3 are the respective sketches of visualizations for these questions:

For (1), we thought a dot plot would be a nice visualization because each line of dots clearly outlines a country and each dot will represent a movie and the value of the dot would be the movie's box office in that country. In addition, by color-coding the dots based on whether or not there is an actor from that country in the movie, we can see visually where the movies would lie on the plot and compare between movies with actors from the country and those without.

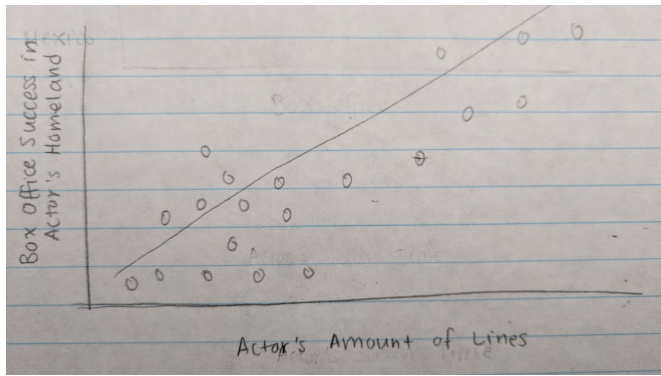
For (2), we suspect that if an actor from country A has a lot of lines in movie X, then the movie would perform much better in country A than another actor from country A who did not have as many lines in a movie. Thus, we are looking to see if there is any correlation that would confirm or even dispel our hypothesis, and that is why we opted to use a scatterplot.

Data Collection

As mentioned earlier, much of our time was spent on collecting the necessary data and enough of it to create visualizations.

Resources for Data

We ended up collecting data on: movies, cast, actors, nationalities, scripts of movies, as well as box office performance of movies. As one can imagine, the data came from varied sources, specifically, IMDB, TMDB, Wikidata, Box Office Mojo, and Wikipedia. As for the scripts, we were unable to find one good source that had the scripts we needed. While there are crowd-sourced scripts available online, most of those scripts do not have characters denoted (i.e. what character is speaking the line), which defeats the purpose for us since we will be looking into the number of lines that each actor speaks. As a result, we chose to manually Google for the scripts.



Picture 3. Scatterplot showing correlation between actor's amount of lines in movie and movie's box office success in actor's homeland.

Web Scraping

R and Python were the primary languages we used to scrape the web for the data. First off, we limited our movie scope to be the top 150 internationally grossing movies every year for the past 11 years (2007-2017). Since our focus was on Hollywood movies, we made sure that countries that contributed to the movies included the US (in other words, we wanted to exclude movies that were only made in countries other than the US such as Bollywood movies). We scraped from IMDB and TMDB information about the movies such as main cast, characters, and year it was released.

Once we have the main cast, we took that list to Wikipedia and Wikidata and filled in information about each cast member's country of citizenship. If we couldn't find information about the actor's nationality, we subbed it for place of birth instead.

Since we were also able to grab the cast and character information, we used the character information to identify their lines in the script and counted the number of lines. We scraped multiple movie script databases in order to obtain as many movie scripts as possible. Those scripts would be used when addressing the question around correlation between number of lines and box office. Scraping the scripts proved to be extremely difficult because a lot of the movie scripts were unavailable. In addition, a lot of scripts we found were in pdf formats that were hard to process for information.

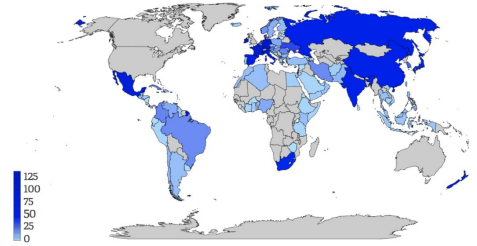
Finally, we not only needed one number for the box office success, but we also needed all the numbers we could get about box office performance in all the countries that the movie was shown in. Thus, we scraped Box Office Mojo to find those numbers.

All in all, we were able to collect data for 1650 movies, 22,018 actors, 60,747 box office reports for the movies, and 391 scripts.

RESULTS

Country	Number of Actors
United Kingdom	2264
Canada	826
Australia	367

World Map of Hollywood Foreign Actors (excluding top 3 countries)



Picture 4. Heat map showing the countries with the most actors/actresses in Hollywood films

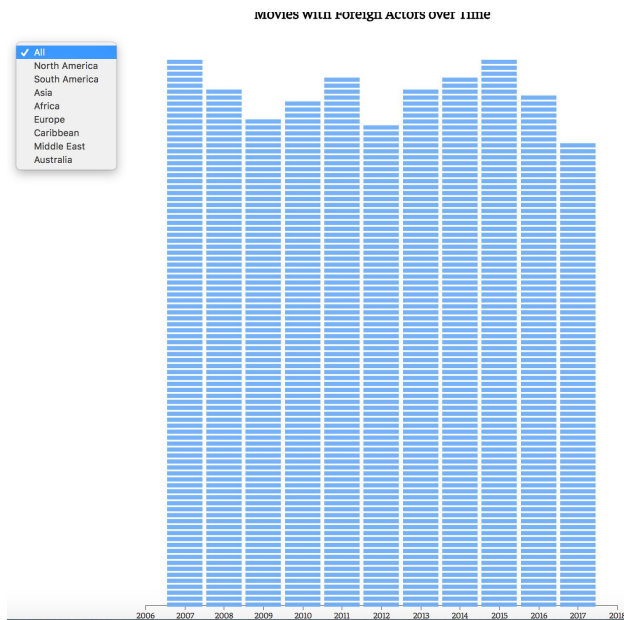
Where do Most Foreign Actors Come from?

We answered this questions using a heat map to show which foreign countries had the most actors in top-grossing Hollywood movies. At first, because an overwhelming number of foreign actors were from English-speaking countries such as the United Kingdom, Canada and Australia, the heat map displayed the other countries as having very similar numbers of actors. The color of those three countries were very similar shades of dark blue while the color of the rest of the countries were the same shade of a much lighter blue. As a result, we decided to remove those countries from the map in order to allow the user to better differentiate and the number of actors from other non-English speaking countries. We ended up listing the number of actors in Hollywood from those top 3 countries as a table above chart.

Is there any Change in the Number of Foreign Actors Over Time?

In order to answer this question, we created a stacked histogram where the x-axis was a year between 2007 and 2017. Because we were counting the number of movies and each movie is it's own discrete unit, we decided to have each bar in the histogram be made of stacked rectangles, where each rectangle represented a movie from a specific year that contained a foreign actor. Hovering over each box would allow the use to see the name of the movie and box office total of the movie represented by that rectangle.

Looking at the initial histogram, it seemed that there were no general trends in the number of foreign actors in Hollywood over time. So we decided to further look into the data and see if there were any trends over time by continent. We added a drop down filter that would allow the

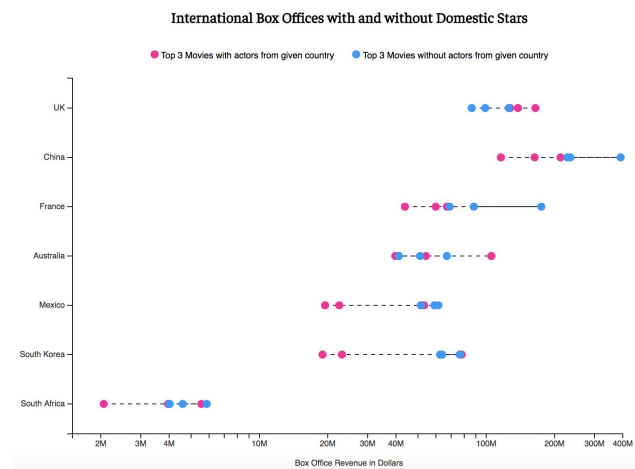


Picture 5. Histogram showing Hollywood movies with foreign actors/actresses over the past 11 years

user to pick a continent and see how the number of foreign actors had changed over time as well. Initially, we found it difficult to change the heights of the histogram when we filtered by continent because each histogram bar was made by stacking rectangles and we would have to recalculate the height of the rectangles after filtering. So we just kept the original histogram and highlighted the boxes representing films from the selected continent. This proved to not be as informative, so we ended up resizing the histogram each time the user selected a new continent.

Is there a Significant Relationship Between the Presence of a Foreign Actor in a Hollywood Movie and its Success in the Foreign Actor's Nation?

We attempted to answer this question by creating a dot plot comparing the top three highest grossing films containing domestic actors and the top three highest grossing films containing only American Hollywood actors in top box offices outside of the US. The x-axis of this plot was the box office of that film. We decided to include at least one country from each continent to be able to compare continental differences as well. After looking a number of graphs online that data similar to the data set that we had for this question, we decided to a dot plot because it was a simple and clear way of answering our question. We made the dots for foreign movies containing domestic and containing only American Hollywood actors noticeably different colors so that users could easily differentiate between which films had foreign actors and which ones only starred American actors. Hovering over each dot would provide the user with the name and box office of the film.



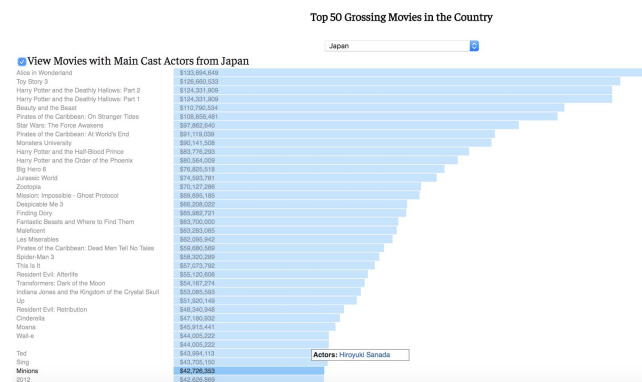
Picture 6. Dot plot comparing box office revenues in countries for movies containing actors from that country and movies with only Hollywood actors

One of our challenges was that the box offices for different countries were on very different scales. For example, the top three box office films in South Africa averaged to about \$4 million, while the top three box office films in China averaged to about \$200 million. Our initial dot plot had a linear x-axis scale. But that made the dots in the plot very clustered together because the dots were either at the lower extreme or higher extreme. It was hard to differentiate between the different dot values. As a result, we decided to use a logarithmic scale in order to allow the user to better see the difference in box office results for the two different types of films.

Are there Thresholds such as an Actor's Screen Time/Amount of Lines that Influence Box Office Success?

To answer our final question, we created a scatter plot showing the relationship between the number of lines an actor had in a movie and the box office in that actor's foreign country. This scatter plot was made to allow the user to filter by country because we believed that it made more sense to compare the number of lines of an actor with the film's box office performance in that actor's country as opposed to the film's general international box office. Hovering over each dot in the scatter plot will provide the user with the film's name and box office, as well as the foreign actor's name, character, and number of lines.

This scatter plot was pretty inconclusive because we were unable to obtain a lot of scripts for the more recent movies. In addition, the scripts of a lot of the movies with some of the largest box offices were also unavailable. In addition, the dots only represent the total number of lines each actor had in a specific film. They did not take into consideration the total number of lines spoken in the film. This may be misleading because certain films have less dialogue in



Picture 7. Bar chart displaying the highest grossing films for each country and which of those films contained actors from the chosen country.

overall. As a result, we feel that the results shown by this scatter plot should be taken with a grain of salt.

Other Interesting Insights - how do the Box-Office Top Grossing Films Differ by Country?

Because we had such a large amount of data, we decided to look into other possibly interesting results. In order to see how box office topping films differed by country, we created a filtered bar chart showing the top box office films

for specific countries. This bar chart was able to show us which movies were the highest grossing in each country.

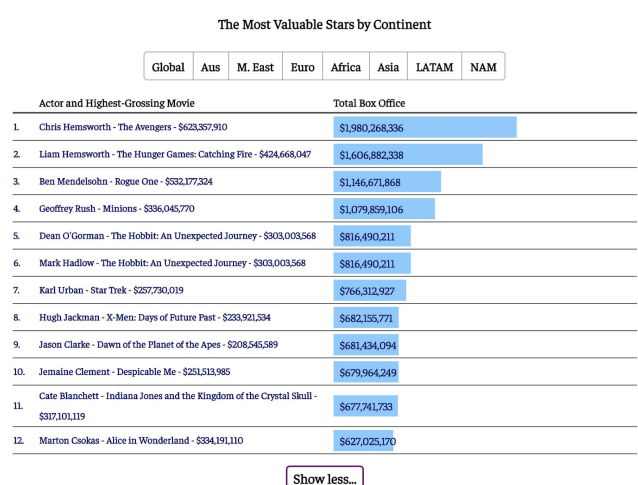
The movies were ranked in order of box office performance, so when users filtered for a different country, they were able to instantly compare the highest or second-highest grossing film from that country with the previous country.

Checking the box for view movies with main cast members from the filtered country would highlight the bars containing actors from the selected country. Hovering over those highlighted bars would then inform the user of which actors in that movie were from the selected country. For some countries, we can see none of their top-grossing films contained actors from that country.

Other Interesting Insights - who are each Continent's Most Valuable Stars?

We also decided to look into who the most valuable stars from each continent. We did so by comparing each actor's overall normalized box office performance. We then created a bar graph showing the top 10 actors/actresses from each continent and how much that actor/actress's movies have made at the box office. We also included in each bar's label the name of the actor's highest-grossing film and how much that film made at the box office.

We had initially hoped to show the top 20 actors/actresses in each continent, but some continents only had around 10 actors from there, so we created a show more button that would allow users to see more actors from certain



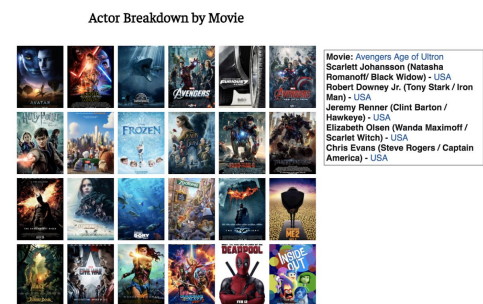
Picture 8. Bar chart displaying the most valuable stars from each continent in terms of average box office revenue

continents. There are also a row of button at the top of the graph that allow the users to filter by continent. We created that row of buttons as filters mainly because our previous filters have all been dropdown menus and we wanted some diversity in our visualizations.

Other Interesting Visualizations - Actor Breakdown by Country

Finally, we decided to create a simple little grid visualization of the 24 most popular and most successful films over the past 11 years. By hovering over each movie poster, users are able to see which countries the stars in that movie are from.

What about your Favorite Movies?



Picture 9. Picture Grid displaying the actor and country breakdown for the top 24 most popular films

Our results can be found on web.stanford.edu/~daeunk.

We used D3.js to generate our visualizations.

DISCUSSION

After showing our product to a wide variety of users including friends, family, and other classmates, we found that overall, users really seemed to enjoy the interactivity of our visualizations and how our webpage walked them through and explained each of the visualizations. In

addition, users seemed to appreciate the variety of visualizations that we included in our final product. We also found that mouseover effects were a clean and efficient way of including necessary information that would have made the visualizations too messy and crowded if we including them any other way.

World Map of Hollywood Foreign Stars

Overall, users did not seem surprised that most foreign actors were from English-speaking countries such as the United Kingdom and Canada. Their first reactions to this visualization were comments such as, “Oh that makes sense, because the UK, Canada, and Australia all speak English,” and “After the English-speaking countries, most foreign actors come from Asia, other parts of Europe, then South America. Maybe the number of foreign actors depends on how developed each region?” The user’s reactions demonstrated that our heat map was a useful tool in representing which countries had the most foreign actors in Hollywood and most users did not require an explanation about this visualization at all.

Box Office numbers by country

As mentioned previously, users were very interested in the interactivity of our visualizations. Users were particularly interested in the actors and top- grossing movies from their home countries would often immediately select their country or continent from the drop down lists of our visualizations. Through these drop down menus and the mouseover effects we placed in most of our visualizations, users were able to learn new information about actors and the box offices of different countries that they had no idea about before.

International Box Offices with and without Foreign Stars

This was probably the most confusing visualization for users from our webpage. When we showed a lot of users our product, we had not yet included a legend for this plot. As a result, most users did not really understand what the dot plot was showing. We ended up having to explain to them what was going on in the chart. Because the aim of our webpage was to provide a walkthrough of our research topic that did not require any outside explanations, we realized that adding a legend was necessary. After adding the legend and some more explanations in the text of our webpage, users were much more clear on what the dot plot was trying to show them.

FUTURE WORK

While completing this project, even though we were able to gather a large amount of data, there are still other kinds of data that we hope to be able to obtain more of. In the future, we would hope to be able to look more into movie scripts to be able to analyze them for content and themes. For example, do actors from certain countries tend to play specific kinds of roles and star in movies with particular themes? We would also be interested in uncovering

differences and insights in gender and racial representation from the foreign actors subset.

Furthermore, we also hope to be able to give more context to the data in terms of what is happening in the real world. For example, have relationships (economic or political), between the US and certain countries influenced Hollywood’s decision to cast foreign actors and movies’ performance in foreign box offices.

Visually, we would hope to give users the ability to search for and learn more about specific movies and more interactivity. In addition, creating animations to display the changes of foreign actors and films over time may be a more interesting way to represent parts of our data.

REFERENCES

1. “Not all American films travel easily. It takes expertise to connect them with foreign audiences.” Los Angeles Times.
<http://www.latimes.com/entertainment/movies/la-c-a-mn-selling-films-abroad-20160417-story.html>
2. “Hollywood blackout: China is giving domestically made films a boost.” CNBC.
<https://www.cnbc.com/2017/08/03/hollywood-blackout-china-is-giving-domestically-made-films-a-boost.html>
3. The Hollywood Insider.
<http://www.informationisbeautiful.net/visualizations/the-hollywood-insider/>
4. “Film Dialogue.” The Pudding.
<https://pudding.cool/2017/03/film-dialogue/>
5. “Are Pop Lyrics Getting More Repetitive”. The Pudding.
<https://pudding.cool/2017/05/song-repetition/>